

**SYSTEM AND METHOD FOR SWITCHING DATA STORAGE
DEVICES AND MULTIPLE PROCESSORS**

5 Inventors: Jack P. Wade, La Jolla, CA.
 Joel Brown, Encinitas, CA

FIELD OF THE INVENTION

10 The present invention relates to a system which provides for switching of the input output (I/O) channel of a storage device between at least two distinct computers.

BACKGROUND

 Highly available computing is the process of designing a computer network so that
15 system operations can continue to operate even with the malfunction or other unexpected interruption to a component of a computing network. Such systems are utilized in situations that demand a high degree of reliability. The goal of highly available networks is to provide duplicate network components to reduce the risk of a single point of failure. In the event of a component failure, duplicate or backup components can take over the role of a failed
20 component, where a component is a general term which can include devices such as a networking switch, a storage device, a computer, or any additional device that may connect to computer network. There are a variety of possible configurations for highly available systems and typically the more effective configurations provide for duplication of components.

25 One example of a prior system 100 is shown in Fig. 1. This system includes two distinct computer systems, computer 102 and computer 104, each with an independent CPU, RAM, I/O bus(es) for user input devices (e.g. keyboard and mouse), and I/O bus(es) for external system devices (e.g. network switches, printers, storage devices, etc.). The system 100 includes an external switch 106 that connects input/output (I/O) ports of each
30 computer system to two separate docking bases, 112 and 114. Each docking base in turn

connects to external removable storage modules 108 and 110. These storage modules are typically individual hard disk drives.

The actual implementation of these systems can be accomplished in a number of different ways. For example, some systems are configured in a system rack device which holds a number of separate boxes, and each box could correspond to a different computer or associated computer device such as storage systems or networking components. Each computer would be able to operate independently and would have its own CPU, power supply, RAM, etc. Storage devices in the rack could take different forms. One such device is a storage module chassis which is designed to hold removable storage disk drives. In environments where security is of concern removable storage modules can be provided where the storage modules are designed to be easily removed and inserted into receptacles of the storage module chassis. Issued U.S. Patent No. 5,126,890, and issued U.S. Patent No. 5,280,398 discuss different aspects of removable disk drive storage modules, and each of these references is incorporated herein by reference in its entirety. Both of these patents are assigned to the same assignee as the present patent application.

In addition to the above patents describing aspects of removable disk drives, U.S. Patent No. 5,552,776 also discusses aspects of removable disk drives, and also describes systems and methods related to providing for security by controlling access between different computers and storage modules. The US Patent No. 5,552,776 is also assigned to the assignee of the present patent application, and is incorporated herein by reference in its entirety.

System 100 of Figure 1 contains removable storage modules 108 and 110. The system 100 allows for each removable storage module to be inserted into a receptacle in a storage module chassis. Each receptacle in the storage module chassis provides a docking base with a connector for receiving a connector from the removable storage module. For example storage module 108 is shown as being coupled with docking base 112 of a storage module chassis, and storage module 110 is coupled with docking base 114 of a storage module chassis. The I/O channel of the storage module 108 is coupled through the docking base 112 to a switch which is external to the docking base and the storage module. The I/O channel of the storage module 110 is coupled through the docking base 114 to the

switch 106. The switch 106 is controlled to provide computers 102 and 104 access to the different storage modules.

This approach of providing an external switch 106 adds potential compatibility and interoperability problems. Such issues increase the complexity and cost of the system, reduce the reliability or uptime of the system, and introduced control issues. For example, if a particular storage module is not working with a computer, then the failure could be in either the storage module or the switch. Further, the central switch acts as a single point of failure. If the switch fails it is likely that the computers may not have access to any of the storage modules.

The above described system is just one example of creating a highly available system utilizing an external switch. Another example of a prior system, is one that utilizes disk storage modules which have two I/O channels and two I/O ports. An existing such storage system is the Fiber Channel (FC) interface drive provided by Seagate Technology LLC. In systems where these drives are configured for RAID operation a single or dual redundant RAID controller can access either port of the drives by means of a hub or a switch inserted in the FC loop between the drives and the controllers. With a hub, there is no switching as all drive ports are seen by the controllers. It is up to the RAID controller programming to arbitrate which computer owns each drive. If a drive port goes bad the RAID controller and the computer can continue using the other port. In a dual RAID controller mode, if one of the RAID controllers fails, the other controller can take ownership the drives. The hub or switch provides the connectivity and the RAID controllers provide the switching, redundancy and failover intelligence. Hubs or switches on the computer channels are required to do failover transparent to the host.

At the network level, between the RAID box and the host computers, FC switches can provide switching and multi-path redundancy. There is usually some storage area network (SAN) control mechanism that involves firmware or software on the host computer, the switches and the RAID controllers. On the RAID controller there is a method called SAN masking, which controls which host computers can have assess to each RAID set. Switches can be zoned to partition traffic and control access. SANs can be very complicated and often have interoperability problems between all of the pieces of the systems. At the host level, multipath software can reroute traffic through a redundant connection to the Raid box.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a diagram illustrating a prior art system.

Fig. 2 is a diagram illustrating an embodiment of the present invention.

5 Fig. 3 is a diagram illustrating an embodiment of the present invention.

Fig. 4 is a diagram illustrating an embodiment of the present invention.

Fig. 5 is a diagram illustrating a storage module chassis and other elements of an embodiment of the invention herein.

Fig. 6 illustrates a removable storage module being inserted into a docking base of
10 storage module chassis.

DETAILED DESCRIPTION

Fig. 2 shows a system 200 of an embodiment of the present invention. The system 200 has two computers 202 and 204 connecting to a docking base backplane 242 of a docking base 240 via multiple I/O ports. Computer 202 connects to the docking base
15 backplane 242 through a Serial ATA (SATA) I/O port 212 coupled to a docking base SATA port 222, and a Universal Serial Bus (USB) I/O port 214 coupled to the docking base USB I/O port 224. Also included is a second computer 204 having a SATA I/O port 216 coupled to the docking base SATA I/O port 226, and a USB I/O port 218 coupled to the docking base
20 USB I/O port 228. In addition, to the connections shown in Figure 2, there would also be I²C connections between computer 202 and docking base back plane 242, and between computer 204 and backplane 242, which are not shown. These I²C connections could operate to couple a PCI RAID Controller within each computer 202 and 204 which implements industry standard SES (Storage Enclosure Services) protocol, in a situation
25 where the removable storage module contains multiple disk drives, operating in a RAID configuration. These I²C connections allow for communication between the raid controllers and a PIC Microcontroller on the docking base backplane 242.

The removable storage system 260 includes two physical units: the docking base unit 240 and the removable storage module 250. Both the docking base unit 240 and the
30 removable storage module 250 can include a housing which can be formed of metal, plastic or another suitable material, and inside the housing are other elements such as backplanes

and components mounted to the backplanes which are discussed in more detail below. Two important elements of the docking base unit include: the docking base backplane 242, and a connector 246 in one embodiment is a 100 pin docking connector. The docking base backplane 242 has I/O connectors for power, Serial ATA interfaces, USB interfaces, and I2C interfaces that connect to the motherboards of computers 202 and 204. The 100 pin docking connector 246 is physically joined to the docking base unit 240 and its backplane 242, and is used to connect the removable storage module 250 to the docking base 240 and the docking base backplane 242 base unit 240 both physically and electronically. The docking base backplane 242 contains a mechanical guiding mechanism to insure proper mating of the connector of the removable storage module 250 with the 100 pin docking connection 246. The docking base backplane 242 also provides a grounding mechanism, in the form of conductive gaskets, to the removable storage module 250 in order to reduce electromagnetic interference. The removable storage module 250 is a device that includes a housing in which multiple disk drives can be mounted. The ability to insert and remove the removable storage module 250 from the docking base unit 240, allows for enhanced security and protection of information stored on the disk drives.

Fig. 3 provides additional details for a system 300 of an embodiment herein. The removable storage module 350 is docked by connecting the docking connector 346 of the removable storage module 350 with the docking connector 347 of the docking base 340 which results in the removable storage module 350 being electrically and physically connected with the docking base 340 and the docking base backplane 342. This mating of the connectors provides power and communication signals for the elements of the removable disk drive module 350. The docking pin connectors of Fig. 3 are shown as 100 pin connectors, but depending on specific implementations, it may be desirable to use a 120 pin connector, or a connector with fewer numbers of pins. In some circumstances providing a higher number of pins can be advantageous because it would allow for providing power and communications with more elements in the removable storage module 350. For example, in one embodiment which is contemplated for used in a cold weather environment it maybe desirable to provide a heating unit in the removable module.

The housing 351 of the removable storage module 350 can include mounting brackets, shock absorbers, or other means for securing hard disk drives 390-396 inside of the

housing. The docking base backplane 342 provides the physical interface between the removable storage module 350 and the corresponding computers 302 and 204. The docking base backplane 342 includes USB, SATA, I²C, and power interfaces. Inside the docking base unit resides a microcontroller 330 which is mounted to the docking base backplane 342.

5 The microcontroller 330 that determines which I/O channel (A or B) will access the individual disk drives located within the removable storage module 350. In one embodiment the microcontroller monitors the two computers 302 and 304 via the USB interfaces 322 and 326 and determines whether the I/O channel will be set to A or B. In one embodiment, the microcontroller 330 will only change the channel if there is a malfunction in the currently
10 active computer system. Ports 315 and 317 provide for I2C signals between computer 302 and the microcontroller 330, and ports 319 and 321 provide for I2C communication between computer 304 and the microcontroller 330.

In one embodiment the removable storage module 350 includes four disk drives 390-396 and a removable storage module backplane 353. Mounted to the removable storage
15 module backplane 353 and associated with each disk drive is a switch 372-378 that connects to the I/O channels 382-388 of the disk drives 390-396. The disk drives could be mounted to a separate backplane, or they could be mounted to the backplane 353. One suitable switch is a serial ATA Failover Switch (VSC7175) available from Vitesse Semiconductor Corp. located in Camarillo, California. The I/O channels can be set to either A or B where each
20 letter corresponds to one of the attached computers 302 and 304. The microcontroller 330 will set the channel of the switch via the control line 371 which is shown as being coupled to switch 372, but additional connections would be provided to each of the switches. In one embodiment, the operation is such that all of the switches for the I/O channels, 382-388 for each of the drives, are coupled via the switches to either the A channel or the B channel, and
25 all of the drives will be connected to the same channel at a particular time. For example, if computer 302 is currently active but suddenly fails, the microcontroller 330 will notify each switch 372-378 to change from channel A to channel B, where channel A corresponds to computer 302 and channel B corresponds to computer 304. The data that is transmitted along I/O channels 382-388 will then switch from channel A to channel B. Under these
30 circumstances, computer 304 will then be the primary computer and all I/O to the disk drives will be controlled by computer 304. As shown the I/O channel for each of the drives is

coupled to a first port of the switches. Each of the switches has a second port which corresponds to an A channel, and the A channel is routed through the docking base backplane to the computer 302. Each of the switches also has a third port which corresponds to a B channel, and the B channel is routed through the docking base backplane to the
5 computer 304.

The microcontroller 330 has several additional functions besides controlling the I/O channel. The microcontroller 330 also monitors the temperature of the system through the temperature sensor 352 located inside the removable storage module 350. Further, data in the EEPROM memory chip 354 is accessed by the microcontroller and can provide
10 additional security, as described in the US Patent No. 5,552,766. The microcontroller 330 also transmits the status of the system to the LED connection 356. The information is then displayed to users of the storage system via a LED display 358. The LED displays such information as disk drive activity, current operational channel, and warning indicators for environmental conditions. In one embodiment, the microcontroller 330 sends and receives
15 communications among the temperature sensor 352 and tracker EEPROM memory 354 via an I²C bus, and LED connection 356 via parallel port bits.

In a preferred embodiment, disk drives 390-396 could be in a RAID configuration to provide enhanced reliability and data recovery should one of the disk drives fail. RAID operation for a group of disk drives is generally widely known, and is a concept first defined
20 by David A. Patterson, Garth Gibson and Randy H. Katz of the University of California, Berkeley in 1987. At its most basic level RAID operation uses an a disk array of a number of small, inexpensive disk drives to exceed the performance of a single, large, expensive disk drive. In addition, since RAIDs use a number of small drives, features can be added to protect against the loss of data when a single drive fails. This redundancy is why Raids have
25 become so popular in high-availability applications. RAID is an acronym for Redundant Array of Independent Disks. There are six levels of RAID: level 0 - level 5. Each level supports a different storage layout scheme on the disk drives, from mirroring to parity striping.

The disk drives or the removable storage module could also support other
30 configurations. Although not shown, the storage module chassis can provide a controller for controlling aspects of the operation of the disk drives. In some implementations the

removable storage module 350 could contain more or fewer than 4 drives per module. For higher numbers of drives per removable storage module it may be necessary to use smaller form factor disk drives. For example, individual 2.5" disk drives, which are now commonly used in laptops, could be utilized in the removable storage module described herein. One
5 suitable 2.5" SATA drive is the Fujitsu Model No. MHT2080AH. Other possible drives can also be used, such as Serial Attached SCSI (SAS) drives. The same physical connector and electrical interfaces can be used for both SATA and SAS drives. Disk drives utilizing high-speed serial communication I/O channels are utilized in one embodiment of the invention. These drives can utilize ATA and SCSI protocols with a serial data interface, instead of a
10 parallel interface. Utilizing a serial interface allows fewer transmission lines per I/O channel. Thus, multiple I/O channels can be provided between a docking base connector and the different CPUs, where in the past the number of lines required for parallel interfaces, frequently made it impractical to provide for multiple I/O channels.

To provide for control of the operation disk drives, for example in a RAID
15 configuration, a controller is required to coordinate the operation of the drives. This controller could be incorporated into the storage system 360, or it could be located external to the storage system 360 in computers 302 and 304. Where the controller is incorporated in the storage system 360, it could be implemented with a PCI card inserted into a PCI slot which is connected with the docking base backplane 342, where such an implementation
20 would require that the docking base backplane contain a motherboard or some sort of computer system to support the PCI slots. Where controllers are external to the storage system 360, the controllers can be either hardware or software based and be either incorporated into, or controlled by computers 302 and 304. Also referenced in Fig. 3 is the hot-swappable controller 380. This controller 380 enables the removable storage module 350
25 to be inserted into and removed from the docking base without requiring recycling of the power. This enables efficient and safe insertion and removal of the storage module 350.

Fig. 4 shows additional details of an embodiment of system 400 herein. In the system 400 a docking base unit 402 is provided. This docking base unit 402 could be one of a number of docking base units which are provided in a storage module chasses, which is
30 shown in more detail in Fig. 5. The docking base unit 402 includes a docking base backplane 404, which can be a printed circuit board to which other components of the

